DOCUMENT RESUME

ED 448 206 TM 032 236

AUTHOR Onwuegbuzie, Anthony J.
TITLE Effect Sizes in Qualitative Research.
PUB DATE 2000-11-21
NOTE 34p.; Paper presented at the Annual Meeting of the Association for the Advancement of Educational Research (AAER) (Ponte Vedra, FL, November 2000).
PUB TYPE Opinion Papers (120) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Research; *Effect Size; *Qualitative Research; *Research Methodology

ABSTRACT

The American Psychological Association Task Force recommended that researchers always report and interpret effect sizes for quantitative data. However, no such recommendation was made for qualitative data. The first objective of this paper is to provide a rationale for reporting and interpreting effect sizes in qualitative research. Arguments are presented that effect sizes enhance the process of "verstehen/hermeneutics" advocated by interpretive researchers. The second objective of the paper is to provide a typology of effect sizes in qualitative research. Examples illustrating various applications of effect sizes are given. For instance, when conducting topological analyses, qualitative analysts only identify emergent themes, yet these themes can be "quantitized" to determine the hierarchical structure of emergent themes. The final objective of this discussion is to illustrate how inferential statistics can be used in qualitative data analysis. This can be accomplished by treating words spoken by individuals or observations emerging from a particular setting as sample units of data that represent the total number of words or observations existing from that sample member or context. Heuristic examples are provided to demonstrate how inferential statistics can be used to provide more complex levels of "verstehen" than is presently found in qualitative research. (Contains 2 figures and 50 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM

ED 448 206

Running head: EFFECT SIZES IN QUALITATIVE RESEARCH

Effect Sizes in Qualitative Research

Anthony J. Onwuegbuzie

Valdosta State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A.J. Onwuegbuzie

TM032236

Paper presented at the annual meeting of the Association for the Advancement of Educational Research (AAER), Ponte Vedra, Florida, November, 21, 2000.

Abstract

The APA Task Force recommended that researchers always report and interpret effect sizes for *quantitative* data. However, no such recommendation was made *for qualitative* data. Thus, the first objective of the present paper is to provide a rationale for reporting and interpreting effect sizes in qualitative research. Arguments are presented that effect sizes enhance the process of *verstehen/hermeneutics* advocated by *interpretive* researchers. The second objective of this paper is to provide a typology of effect sizes in qualitative research. Examples are given illustrating various applications of effect sizes. For instance, when conducting typological analyses, qualitative analysts only identify emergent themes; yet, these themes can be quantitized to ascertain the hierarchical structure of emergent themes. The final objective is to illustrate how inferential statistics can be utilized in qualitative data analyses. This can be accomplished by treating words arising from individuals, or observations emerging from a particular setting, as sample units of data that represent the total number of words/observations existing from that sample member/context. Heuristic examples are provided to demonstrate how inferential statistics can be used to provide more complex levels of *verstehen* than is presently undertaken in qualitative research.

Effect Sizes in Qualitative Research

One of the most common errors in quantitative analyses involves the incorrect interpretation of statistical significance and the related failure to report and to interpret effect sizes (i.e., variance-accounted for effect sizes or standardized mean differences) (e.g., Onwuegbuzie, 1999; Onwuegbuzie & Daniel, 2000, in press; Thompson, 1998a, 1998b, 1999; Thompson & Daniel, 1996). This error often leads to under-interpretation of associated *p*-values when sample sizes are small and the corresponding effect sizes are large, and an over-interpretation of *p*-values when sample sizes are large and effect sizes are small (e.g., Daniel, 1998a, 1998b; Onwuegbuzie & Daniel, 2000, in press; Thompson, 1998a, 1998b). Apparently, many analysts operate under the false illusion that their *p*-values (a) test result importance, (b) test result replicability, and (c) evaluate effect magnitude (Thompson, 1998). This is despite the fact that the literature is replete with information about the importance of effect size reporting. In fact, recently, the American Psychological Association (APA) Board of Scientific Affairs, who convened a committee called the Task Force on Statistical Inference, recommended in no uncertain terms, that effect size estimates always be presented when reporting *p*-values (Wilkinson & the Task Force on Statistical Inference, 1999).

According to the APA Task Force, researchers should "always present effect sizes for primary outcomes...[and]...reporting and interpreting effect sizes...is essential to good research" (Wilkinson & the Task Force on Statistical Inference, 1999, pp. 10-11). However, as indicated by the title of their report (i.e., "Statistical Methods in Psychology Journals: Guidelines and Explanations"), it is clear that these recommendations pertain only to

quantitative data. That is, no recommendation was made to report and to interpret effect sizes when analyzing *qualitative* data. Yet, there are many instances in which effect sizes would provide a thicker description of underlying qualitative data. Indeed, it appears that the non-use of effect sizes by qualitative researchers stems, at least in part, from educational researchers associating effect sizes with the quantitative paradigm. As such, many qualitative researchers believe that use of effect sizes will result in the quantitative paradigm being the standard against which qualitative research will be measured. Yet, ironically, use of effect sizes actually *qualitizes* empirical data by helping data analysts to determine whether an observed effect is small, medium, large, or the like--decisions which represent *qualitative* categorizations.

Thus, the first purpose of the present paper is to provide a rationale for reporting and interpreting effect sizes in qualitative research. The second objective of this article is to provide a typology of effect sizes in qualitative research. The final purpose is to illustrate how inferential statistics can be utilized in qualitative data analyses.

Toward a Framework for Unifying Quantitative and Qualitative Research Paradigms

Much of the quantitative-qualitative debate has involved the practice of polemics, which has tended to obfuscate rather than to clarify, and to divide rather than to unite educational researchers. Indeed, as Miles and Huberman (1984, p. 21) stated, "epistemological purity doesn't get research done." On the other hand, epistemological ecumenism allows researchers to re-frame how research paradigms should be viewed. As noted by Newman and Benz (1998), rather than representing a dichotomy, positivist and non-positivist philosophies lie on an epistemological continuum. Indeed, all the various

dichotomies that are used to distinguish quantitative and qualitative paradigms should be re-conceptualized as lying on continua. These include realism versus idealism, foundational versus antifoundational, objective versus subjective, personal versus impersonal, deductive reasoning versus inductive reasoning, generalization versus uniqueness, logistic versus dialectic, rationalism versus naturalism, specific versus holistic, causal versus acausal, and correspondence versus coherence. Such a re-framing allows researchers to focus more on research strategies rather than on paradigmatic issues.

According to Onwuegbuzie and Teddlie (in press), one way of re-framing research in the social and behavioral sciences in general and the field of education in particular is to de-emphasize the terms quantitative and qualitative research and, instead, sub-divide research into *exploratory* and *confirmatory* methods. Such a re-conceptualization unites quantitative and qualitative data collection and data analytical procedures under the same framework. In Onwuegbuzie and Teddlie's (in press) model, quantitative data analysis techniques that are labeled as exploratory include descriptive statistics, exploratory factor analysis, and cluster analysis, whereas exploratory qualitative data analysis involves the traditional thematic analyses. With regard to confirmatory methods, quantitative data-analytical techniques comprise the array of inferential statistics, whereas qualitative data-analytic methods involve *confirmatory thematic analyses*, in which replication qualitative studies are conducted to assess the replicability of previous emergent themes (i.e., research driven) or to test an extant theory (i.e., theory driven), when appropriate.

Such a framework promotes the development of bi-researchers, a term coined by Onwuegbuzie (2000b) to denote researchers who routinely utilize both quantitative and

qualitative research techniques. Indeed, Onwuegbuzie (2000b) goes so far as to recommend that quantitative and qualitative research courses be re-designed as courses in exploratory and confirmatory techniques that teach quantitative and qualitative methodologies within each course, either simultaneously or in a sequential manner. The idea of qualitative and quantitative research faculty team-teaching a course would be truly innovative. In any case, such courses would send a strong message to students that applied quantitative and qualitative research, for the most part, have the same goal, namely to understand phenomena one study at a time. Consequently, students enrolled in these courses will view research as a holistic endeavor, as recommended by Newman and Benz (1998). Additionally, these courses would allow students to focus on the similarities of quantitative and qualitative research, rather than on the differences, with a similarity being the importance of interpreting findings in their proper context via the use of effect sizes. It is within this framework of exploratory and confirmatory data analysis that the following discussion of effect sizes in qualitative research takes place.

Exploratory Qualitative Analyses: A Typology of Effect Sizes

Just as it could be argued that all data are essentially qualitative (Berg, 1989) inasmuch as they represent an attempt to capture a raw experience, so it could be contended that all data can be expressed dichotomously, that is, as a binary variable (i.e., "1" vs. "0") (Sechrest & Sidana, 1995). With respect to the latter, as noted by Sechrest and Sidana (1995, p.79), "every qualitative assertion--'the sky is blue'--can be expressed in binary quantitative form." Moreover, every theme that emerges from the data can be classified as either occurring or not occurring. This ability to *binarize* (i.e., dichotomize)

allows effect sizes to be reported for qualitative data.

When conducting thematic analyses, qualitative analysts typically only classify and describe emergent themes. Although identification of themes represents an extremely powerful way of data reduction (Miles & Huberman, 1994), even more information from these themes often can be extracted. Specifically, whether themes are theory driven, prior data/research driven, or inductive, on every occasion, these themes can be quantitized (i.e., quantified) by determining the frequency of occurrence (e.g., most/least dominant theme) and/or intensity of each identified theme. Indeed, as noted by Sechrest and Sidani (1995, p. 79), "qualitative researchers regularly use terms like 'many,' 'most,' 'frequently,' 'several,' 'never,' and so on. These terms are fundamentally quantitative." In fact, by obtaining counts, qualitative researchers can quantitize such terms. This indicates that numbers and words co-exist in virtually every research setting. We, as researchers, can choose to collect only one type of data and ignore the other type (e.g., words) and thus use only one lens, or we can collect both types of data, utilizing bi-focal lenses. Indeed, it could be argued that the only important difference between quantitative and qualitative data is that the former represent more *empirical precision*, whereas the latter represent more *descriptive precision*.

The frequency of emergent themes (i.e., frequency effect size) can be determined by first binarizing themes. Specifically, for each participant in the study, a score of "1" is given for a theme if it represents a significant statement or observation pertaining to that individual; otherwise, a score of "0" is given for that theme. That is, for each sample member, each theme is binarized either to a score of "1" or a "0," depending on whether

it is represented by that individual. This binarization leads to the formation of an *inter-respondent matrix* (i.e., *participant x theme matrix*) and an *intra-respondent matrix* (i.e., *unit x theme matrix*). Both matrices contain a combination of 0s and 1s.

The inter-respondent matrix indicates which individuals contributes to each theme that emerges, whereas the intra-respondent matrix identifies which units (i.e., significant statements or observations) contribute to each theme that emerges. For qualitative studies that involve more than one participant, both the inter-respondent matrix and the intra-respondent matrix can be utilized; for qualitative studies that involve exclusively one participant, the intra-respondent matrix comes into play. Although binarizing themes can be criticized as an oversimplification of emergent themes that does not capture the complexity of the meaning conveyed by the unit, as stated by Sechrest and Sidani (1995, p. 79), the individual making the statement or action "would have to have shared understanding of all those additional meanings, in which case the binary code would include them all, or else the statement would have to be accompanied by a set of additional descriptors/modifiers that could themselves be coded."

Moreover, the justification for binarizing themes is no less strong as for measuring cognitive performance. Indeed, when measures of academic achievement are administered, responses to standardized test items typically are reduced to an inter-respondent matrix under the assumption that the binarization leads to an approximation of test takers' ability. These inter-respondent matrices stemming from test scores are then used to conduct an array of descriptive statistical techniques (e.g., means, percentile ranks) that inform educational policy. in any case, the goal of binarizing themes is not to

replace the description of the themes, but to facilitate identification of effect size indices that would supplement these descriptions. The binarizing of themes allows the computation of two types of effect sizes, which, hereafter, will be termed *manifest effect sizes* and *latent effect sizes*.

Manifest effect sizes. Manifest effect sizes represent effect sizes that pertain to observable content. This class of effect sizes represents specific counts of significant statements (e.g., words, phrases, sentences, paragraphs, pages) or observations analyzed that underlie emergent themes.

Frequency (manifest) effect sizes are obtained by calculating the frequency of each theme from the inter-respondent matrix. These frequencies can then be converted to percentages in order to determine the prevalence rate of each theme. *Intensity (manifest) effect sizes*, which are determined via the intra-respondent matrix, represent the frequency of each significant statement within each theme. As before, intensity effect sizes can be converted to percentages.

Adjusted effect sizes also can be computed in which the frequency and intensity of themes are adjusted for the time sequence and length of the unit of analysis (e.g., observation, interview, text). For example, with respect to the latter (i.e., length of unit analysis), the number of times that a theme emerges could be divided by the number of (transcribed) words/sentences/paragraphs/pages analyzed. Such adjusted effect sizes help to reduce bias in the data sampled. Additionally, *a fixed-interval effect size index* could be estimated via the inter-respondent matrix or the intra-respondent matrix, in which the frequency (i.e., *fixed-interval frequency effect size)* and intensity (i.e., *fixed-interval intensity*

effect size) of themes are determined as they occur within a specific period of time. For example, a researcher could investigate how many times a word is used in the first 10 minutes of a focus group. Further, a *fixed-ratio effect size index* could be assessed, in which a specific frequency (i.e., *fixed-response frequency effect size)* and intensity (i.e., *fixed-response intensity effect size)* of themes are specified *a priori*, and the amount of time that elapses before these targets are met, if at all, is utilized as an effect size estimate.

Interestingly, an exploratory factor analysis can be undertaken on the inter-respondent matrices and the intra-respondent matrices in order to determine the hierarchical structure of the themes. Factors that emerge from this analysis, which hereafter will be termed *meta-themes*, represent themes at a higher level of abstraction than the original emergent themes. The manner in which the emergent themes cluster within each factor (i.e., meta-theme) facilitates identification of the inter-relationships among the themes. Once the meta-themes have been determined, an *inter-respondent meta-theme matrix* (i.e., *participant x meta-theme matrix*) and an *intra-respondent thematic matrix* (i.e., *unit x meta-theme matrix*) can be constructed comprising a combination of 0s and 1s. These matrices can then be used to determine *frequency (manifest) effect sizes* and *intensity (manifest) effect sizes* for the meta-themes.

Latent effect sizes. Latent effect sizes, the other class of effect sizes, represent effect sizes that pertain to non-observable, underlying aspects of the phenomenon being studied. They are more interpretative than are manifest effect sizes. For example, correlational analyses also could then be performed using the inter-respondent and intra-respondent matrices to determine the relationship among the themes. Correlational

analyses also could be undertaken using the *inter-respondent meta-theme matrix* and the *intra-respondent thematic matrix* to determine the relationship among the meta-themes. The correlation indices contained in these correlation matrices serve as *bivariate latent effect sizes*. Additionally, the exploratory factor analysis undertaken on the inter-respondent matrices and intra-respondent matrices, described above, can be used to compute *variance-explained latent effect sizes*, stemming from the eigenvalues and the proportion of variance explained after rotation (i.e., trace) by each theme.

Finally, the inter-respondent matrix can be used to conduct narrative profile formation (i.e., modal profiles, average profiles, holistic profiles, comparative profiles, and normative profiles; Tashakkori & Teddlie, 1998). For example, the number of average profiles (Tashakkori & Teddlie, 1998) can be determined using an ipsative approach, in which participants' responses to each theme can be interpreted relative to their responses to the other themes (Allport, 1937, 1962, 1966; Block, 1957; Stephenson, 1953) in the following manner: (a) for each participant, the emergent theme scores (i.e., 0 or 1) are ranked such that each scale takes on a value from 1 through *t*, where t represents the number of themes; and (b) the measure of similarity used for the analysis is based on the theme scores ranked from lowest to highest within each profile. An intra-individual correlation matrix is then formed by correlating each pair of profiles, yielding $(n)(n\text{-}1)/2$ Spearman Rho values (where n was the number of respondents). This correlation matrix is then cluster-analyzed in order that individualistic patterns could be characterized for each sample member. Participants having similar profiles are expected to cluster together. The criterion of percentage variation explained by each cluster helps to identify the most

meaningful cluster solution. The formation of average profiles represents the *qualitizing* of previously-quantitized themes (Tashakkori & Teddlie, 1998). This eigenvalues for each cluster-solution are compared to determine the number of interpretable profiles. Each profile can then be compared and contrasted by determining whether, within each theme, the confidence intervals (i.e., standard error bars) overlap, as well as by computing within-theme manifest effect sizes. These within-theme manifest effect sizes involve comparing the average profiles of the clusters within each theme. Standardized mean differences and adjusted/unadjusted variance-accounted-for effect sizes can be utilized as manifest and latent effect size estimates, respectively. Also, each profile group can compared with respect to the selected demographic variables.

Confirmatory Qualitative Analyses: A Typology of Effect Sizes

Historically, in confirmatory studies, whereby inferential analyses prevail, the data collected and analyzed have been quantitative (Tashakkori & Teddlie, 1998). However, inferential statistics also can be utilized in qualitative data analyses, regardless of sample size. Such a treatment of qualitative data is justified by treating words that arise from a person(s), or observations that emerge from a particular setting, as sample units of data that represent the total number of words/observations existing from that sample member/context. Consequently, inferential techniques can be used to generalize words and observations that arise from *persistent observations* and *prolonged engagement* to the *population* of words/observations (i.e., the *truth space*) representing the underlying context (although no generalizations beyond this context is justified), or even to individuals beyond the sample (i.e., the underlying population) if a large enough sample and a careful

sampling design is used.

A common goal of qualitative researchers, especially when interviewing and focus techniques are used, is to capture the voice of the person(s) being studied. Regardless of the number of interviews conducted (i.e., single vs. multiple), the length of each interview, type of interviews (e.g., unstructured, partially structured, semi-structured, structured, totally structured), and format of interviews (e.g., formal vs. informal), words collected represent a mere sample of the interviewee's voice (i.e., truth space). Thus, when conducting thematic analyses, *inferences* are made from the sample of words to the interviewee's truth space. Just as quantitative researchers hope that their sample is representative of the population, qualitative researchers hope that the sample of words is representative of the truth space. However, if the sample of words collected is not representative of the interviewee's total truth space, then the voice sampling error will be large. Consequently, any subsequent analyses of the sample of words will likely lead to untrustworthy findings.

Because, then, inferences are made during qualitative data analyses, an array of statistical techniques, including all those belonging to the *general linear model*, can be utilized to examine trends in the thematic structure. Specifically, for qualitative studies that involve several participants, the antecedent correlates of the emergent themes can be determined via the inter-respondent matrix. For example, a series of Fisher's Exact tests can be used to determine which nominally-measured demographic variables are related to each of the themes. Cramer's *V* statistic can serve as a latent effect size. Further, for demographic variables with two levels (e.g., gender), odds ratios can be utilized as latent

effect sizes. Odds ratios among the meta-themes also can be determined and used to compare prevalence rates among the meta-themes. Alternatively, a canonical correlation analysis can be undertaken to examine simultaneously the relationship between the themes and the demographic variables. Here, for each significant canonical correlation, the canonical correlation and standardized canonical function coefficients and structure coefficients can serve as latent effect sizes. For qualitative studies that involve multiple interviews of one participant, a time series analysis of the themes can be performed.

Heuristic Example

The study from which an example has been selected to illustrate how effect sizes can lead to a thicker, richer description of qualitative data was conducted on 219 preservice teachers attending a large mid-southern university (Witcher, Onwuegbuzie, & Minor, in press). The purpose of this investigation was to determine their perceptions about the characteristics of effective teachers. These preservice teachers were administered a questionnaire asking them to identify, to rank, and to define between 3 and 6 characteristics that they believed excellent teachers possess or demonstrate.

Witcher et al. (in press) conducted what they termed a *sequential mixed-methodological analyses (SMMA)*. This analysis involved utilizing qualitative and quantitative data analytic techniques in a sequential manner, commencing with qualitative analyses, followed by quantitative analyses that built on the qualitative analyses, and then ending with qualitative analyses. The SMMA involved five stages.

Stage 1. The first stage consisted of a phenomenological mode of inquiry (i.e., exploratory stage) to examine the responses of students regarding their perceptions of

characteristics of effective teachers (Goetz & Lecompte, 1984). As noted by the authors, the phenomenological method essentially represents an attempt to understand phenomena from the perspective of those being studied. Thus, the researchers attempted not to form any *a priori* hypotheses (i.e., bracketing) with respect to preservice teachers' perceptions of effective teacher characteristics. Witcher et al. utilized a modification of Colaizzi's (1978) phenomenological analytic methodology, comprising a 5-step method of generating themes, which included *unitizing* the data, *horizonalization* of data, and the *method of constant comparison*; Glaser & Strauss, 1967; Lincoln & Guba, 1985). *Double coding* (Miles & Huberman, 1994) was used for categorization verification in the form of inter-rater reliability.

Stage 2. The second stage of their mixed-methodological analysis involved utilizing descriptive statistics (i.e., exploratory stage) to analyze the hierarchical structure of the emergent themes. In particular, each theme was *binarized*. That is, as described above, for each participant, each theme was quantitized either to a score of "1" or a "0" depending on whether it was represented by that individual. This dichomotization produced an *inter-respondent matrix* (i.e., *participant x theme matrix*) and an *intra-respondent matrix* (i.e., *unit x theme matrix*), which allowed the computation of two types of *manifest effect sizes*. Specifically, the researchers determined the prevalence rate of each theme by calculating the frequency of each theme from the inter-respondent matrix and then converting these frequencies to percentages. These percentages provided a *frequency effect size* measure. Witcher et al. also obtained an *intensity effect size* measure by calculating the proportion of characteristics identified per theme.

Stage 3. The third stage of the mixed-methodological analysis involved the utilization of the inter-respondent matrix to conduct an exploratory factor analysis to ascertain the underlying structure of these themes (i.e., exploratory stage). This factor analysis determined the number of factors (i.e., *meta-themes*) underlying the themes. The *trace*, or proportion of variance explained by each factor after rotation, was utilized as a *latent* effect size for each meta-theme. Further, a manifest effect size was computed for each meta-theme by determining the combined frequency effect size for themes within each meta-theme.

Stage 4. The fourth stage of the mixed-methodological analysis involved the determination of antecedent correlates of the emergent themes that were extracted in Stage 1 and quantitized in Stage 2 (i.e., confirmatory analyses). This phase utilized the inter-respondent matrix to undertake (a) a series of Fisher's Exact tests to determine which background variables were related to each of the themes; and (b) a canonical correlation analysis to examine simultaneously the relationship between the themes and the demographic variables. With respect to the latter, standardized canonical function coefficients and structure coefficients were computed, which served as *inferential-based effect sizes*.

Stage 5. The fifth and final stage of the mixed-methodological analysis involved narrative profile formation. Witcher et al. ascertained the number of average profiles (Tashakkori & Teddlie, 1998) using an ipsative approach in which the preservice teachers' responses to each theme were interpreted relative to their responses to the other themes (Allport, 1937, 1962, 1966; Block, 1957; Stephenson, 1953), using the following steps: (a)

for each participant, the emergent theme scores (i.e., 0 or 1) were ranked such that each scale took on a value from one through six; and (b) the measure of similarity used for the analysis was based on the theme scores ranked from lowest to highest within each profile. An intra-individual correlation matrix was then formed by correlating each pair of profiles, yielding $(n)(n\text{-}1)/2$ Spearman Rho values (where n was the number of respondents). This correlation matrix was then cluster-analyzed such that individualistic patterns could be characterized for each preservice teacher. The formation of average profiles represented the *qualitizing* of previously-quantitized themes (Tashakkori & Teddlie, 1998).

The phenomenological analysis of responses (i.e., Stage 1 and Stage 2) revealed several characteristics that many of the preservice teachers considered to be indicative of effective teaching. In order of endorsement level, Witcher et al. found the following six emergent themes: (a) student-centeredness (79.5%), (b) enthusiasm for teaching (40.2%), (c) ethicalness (38.8%), (d) classroom and behavior management (33.3%), (e) teaching methodology (32.4%), and (f) knowledge of subject (31.5%). Additionally, an examination of the intercorrelations among the six themes, after applying the Bonferroni adjustment (Onwuegbuzie & Daniel, in press), revealed a statistically significant but small relationship between responses to the classroom and behavior management theme and the enthusiasm for teaching theme (i.e., $r = .20$, $p < .003$). However, this was the only statistically significant relationship found by the authors out of the 15 possible relationships among the themes, which suggested that these themes were somewhat independent of one another.

The exploratory factor analysis (Stage 3) revealed that the six themes were

subdivided into the following four meta-themes: classroom atmosphere (comprising the classroom and behavior management and enthusiasm themes), subject and student (comprising the knowledge of subject and student-centeredness themes), ethicalness (comprising the ethicalness theme), and teaching methodology (comprising the teaching methodology theme). The thematic structure is presented in Figure 1. This figure illustrates the relationships among the themes and meta-themes arising from preservice teachers' perceptions of the characteristics of effective teachers.

Insert Figure 1 about here

An examination of the *trace* (i.e., the proportion of variance explained, or eigenvalue, after rotation; Hetzel, 1996) revealed that the *classroom atmosphere* meta-theme explained 20.65% of the total variance, the *subject and student* meta-theme accounted for 19.07% of the variance, the *ethicalness* meta-theme explained 18.26% of the variance, and the *teaching methodology* meta-theme accounted for 16.74% of the variance. These four meta-themes combined explained 74.7% of the total variance. As noted by the investigators, the total proportion of variance represented a *latent effect size*. Witcher et al. also computed manifest effect sizes associated with the four meta-themes (i.e., proportion of characteristics identified per meta-themes) as follows: *classroom atmosphere* (64.8%), *subject and student* (88.6%), *ethicalness* (38.8%), and *teaching methodology* (32.4%). Thus, as they noted, both the latent and manifest effect sizes associated with these meta-themes were moderate to large.

The canonical correlation analysis (Stage 4) revealed that females, college-level juniors, and minority students tended to endorse teacher characteristics that were associated with ethical behavior and teaching methodology to a greater extent than did their counterparts. These subgroups also tended to rate attributes that were associated with knowledge of subject and classroom and behavior management to a lesser degree. Age served as a suppressor variable. The canonical correlation was moderately educationally significant, contributing 19.4% to the shared variance.

Finally, Witcher et al. conducted an ipsative/cluster analyses (Stage 5), which revealed four profiles of students' responses to the six themes. Each of the four emergent profiles represented an average set of responses across each theme. The profiles for the resulting four clusters are reproduced in Figure 2. As can be seen from this diagram, members of Cluster 1 (n = 56) were very likely to endorse the student-centeredness (probability (p) = .84) and enthusiasm for teaching (p = .71) themes. These preservice teachers were moderately likely to endorse the teaching methodology theme (p = .41); however, they were unlikely to endorse the knowledge of subject (p = .30), classroom and behavior management (p = .16), and ethicalness (p = .11) themes.

Individuals in Cluster 2 (n = 51) also highly rated student-centeredness (p = .83). Further, they were very likely to endorse classroom and behavior management (p = .16); however, they were unlikely to cite a characteristic associated with the teaching methodology (p = .27), enthusiasm for teaching (p = .21), ethicalness (p = .18), and knowledge of subject (p = .14) themes. Members of Cluster 3 highly rated student-

centeredness (p = .83) and ethicalness (p = .85). On the other hand, Cluster 3 sample members were unlikely to cite a characteristic pertaining to the enthusiasm for teaching (p = .37), teaching methodology (p = .25), classroom and behavior management (p = .25), and knowledge of subject (p = .22) themes. Finally, preservice teachers in Cluster 4 were highly likely to endorse the student-centeredness theme (p = .74) and knowledge of subject theme (p = .68). They were moderately likely to endorse the ethicalness (p = .40) and teaching methodology (p = .40) themes; however, they were unlikely to endorse the enthusiasm for teaching (p = .32) and classroom and behavior management (p = .30) themes.

Insert Figure 2 about here

Summary

Although the APA Task Force and others (e.g., Onwuegbuzie, 1999; Onwuegbuzie & Daniel, 2000, in press; Thompson, 1998a, 1998b, 1999; Thompson & Daniel, 1996) recommend that effect sizes always be computed and reported in quantitative studies, there is no such recommendation for qualitative research. Yet, there are many instances in which effect sizes would provide a thicker description of underlying qualitative data. Indeed, use of effect sizes actually *qualitizes* empirical data by helping data analysts to determine whether an observed effect is small, medium, large, or the like--decisions which represent *qualitative* categorizations.

Thus, the first purpose of the present paper was to provide a rationale for reporting

and interpreting effect sizes in qualitative research. Arguments were presented that effect sizes enhance the process of *verstehen/hermeneutics* advocated by *interpretive* researchers. A historical background of the quantitative-qualitative debate was discussed. This account included a description of the influential positivist theories of Comte; Dilthey's interpretive/hermeneutical approach to science, which represented the first serious challenge to positivism; and Weber's attempt to synthesize the two research paradigms. It was contended that no one paradigm is a hegemony in educational research. In fact, although quantitative and qualitative research paradigms are distinct, they are somewhat related, inasmuch as at any moment in time, quantitative and qualitative data co-exist for virtually every phenomena of interest to us in the world. Also, the claim by purists that quantitative and qualitative research designs are not compatible was refuted. Moreover, evidence was provided that rejects the assertions of purists on both ends of the epistemological continuum. In so doing, several myths held by these purists were identified. It was noted that the fundamental problem with the position of both sets of purists is that their assumptions are self-refuting.

Moreover, it was argued that recognizing such myths allows one to re-frame how research paradigms should be viewed. It was contended that one way of re-framing research is to de-emphasize the terms quantitative and qualitative research and, instead, sub-divide research into *exploratory* and *confirmatory* methods. Moreover, it was asserted that such a re-conceptualization would unite quantitative and qualitative data collection and data analytical procedures under the same framework.

The second objective of this paper was to provide a typology of effect sizes in

qualitative research. Examples were given illustrating various applications of effect sizes. For instance, it was noted that when conducting typological analyses, qualitative analysts only identify emergent themes; yet, these themes can be quantitized to ascertain the hierarchical structure of emergent themes. An array of manifest effect sizes (i.e., effect sizes pertaining to observable content) and latent effect sizes (i.e., effect sizes pertaining to non-observable, underlying aspects of the phenomenon under observation) were outlined for both exploratory and confirmatory qualitative data analyses.

The third purpose was to illustrate how inferential statistics can be utilized in qualitative data analyses, regardless of sample size. It was argued that this can be accomplished by treating words arising from individuals, or observations emerging from a particular setting, as sample units of data that represent the total number of words/observations existing from that sample member/context. Consequently, inferential techniques can be used to generalize words and observations that arise from *persistent observations* and *prolonged engagement* to the *population* of words/observations (i.e., the *truth space*) representing the underlying context (although no generalizations beyond this context is justified), or even to individuals beyond the sample (i.e., the underlying population) if a large enough sample and a careful sampling design is used. An heuristic example was provided to demonstrate how an array of effect sizes can be generated from qualitative data.

Conclusion

In order to promote the use of effect sizes in qualitative research, both quantitative and qualitative researchers must make a distinction between research method as a

technique (i.e., research design) and research method as a logic of justification (i.e., research paradigm), as well as a distinction between research design and data analysis. In so doing, as detailed in the present essay, the full complement of available research designs and analyses can be employed more holistically. Moreover, computing and reporting effects sizes in qualitative research will assist in bridging the wide gap that presently exists between many quantitative and qualitative researchers. Moreover, effect size analyses in interpretive research will serve as a mode for translating between quantitative and qualitative data. Indeed, as noted by Miles and Huberman (1984), to make qualitative findings available to as many individuals as possible, interpretivists must incorporate a myriad of ways of organizing and presenting them. Thus, effect sizes offer a way of including quantitative researchers in the dialogue when interpreting themes. Finally, the use of effect sizes in qualitative data analysis and interpretation can be used to provide more complex levels of *verstehen* than is presently undertaken in qualitative research.

References

Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Allport, G.W. (1962). The general and the unique in psychological science. *Journal of Personality, 30*, 405-422.

Allport, G.W. (1966). Traits revisited. *American Psychologist, 21*, 1-10.

Aron, R. (1970). The logic of the social sciences. In D. Wrong (Ed.), *Max Weber*. Englewood Cliffs, NJ: Prentice-Hall.

Benton, T. (1977). *Philosophical foundations of the three sociologies*. London: Routledge and Kegan Paul.

Berg, B.L. (1989). *Qualitative research methods for the social sciences*. Boston, MA: Allyn & Bacon.

Block, J. (1957). A comparison between ipsative and normative ratings of personality. *Journal of Abnormal and Social Psychology, 54*, 50-54.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: John Wiley.

Cohen, J. (1997). *The earth is round ($p < .05$)*. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 21-35). Mahwah, NJ: Erlbaum.

Colaizzi, P.F. (1978). Psychological research as the phenomenologist views it. In R. Vaile & M. King (Eds.), *Existential phenomenological alternatives for psychology* (pp.

48-71). New York: Oxford University Press.

Constas, M.A. (1992). Qualitative data analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29,* 253-266.

Creswell, J.W. (1995). *Research design: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Daniel, L.G. (1998a). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for editorial policies of educational journals. *Research in the Schools, 5,* 23-32.

Daniel, L.G. (1998b). The statistical significance controversy is definitely not over: A rejoinder to responses by Thompson, Knapp, and Levin. *Research in the Schools, 5,* 63-65.

Ermarth, M. (1978). Wilhelm Dilthey: *The critique of historical reason.* Chicago, IL: University of Chicago Press.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Goetz, J.P., & Lecompte, M.D. (1984). *Ethnography and the qualitative design in educational research.* New York: Academic Press.

Hammersley, M. (1992). Some reflections on ethnography and validity. *Qualitative Studies in Education, 5*(3), 195-203.

Hanson, N.R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science.* Newbury Park, CA: Sage.

Hetzel, R.D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol 4, pp. 175-206). Greenwich, CT: JAI Press.

Hodges, H. (1944). *Wilhelm Dilthey: An introduction.* London: Routledge and Kegan Paul.

Hodges, H. (1952). *The philosophy of Wilhelm Dilthey.* London: Routledge and Kegan Paul.

Howe, K.R. (1988). Against the quantitative-qualitative incompatability thesis or dogmas die hard. *Educational Researcher, 17*, 10-16.

Hughes, H. (1958). *Consciousness and society.* New York: Knopf.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Miles, M. & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Newman, I., & Benz, C.R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum.* Illinois: Southern Illinois University Press.

Onwuegbuzie, A.J. (1999, September). *Common analytical and interpretational errors in educational research.* Paper presented at the annual meeting of the European Conference on Educational Research (ECER), Lahti, Finland.

Onwuegbuzie, A.J. (2000a, November). *Validity and qualitative research: An oxymoron?* Paper to be presented at the annual meeting of the Association for the Advancement of Educational Research (AAER), Ponte Vedra, Florida.

Onwuegbuzie, A.J. (2000b). *On becoming a Bi-Researcher: The importance of*

combining quantitative and qualitative research methodologies. Unpublished manuscript, Valdosta State University, Valdosta, GA.

Onwuegbuzie, A.J. (2000c, November). *Expanding the Framework of internal and external validity in quantitative research*. Paper presented at the annual meeting of the Association for the Advancement of Educational Research (AAER), Ponte Vedra, Florida.

Onwuegbuzie, A.J., & Daniel, L.G. (2000, April). *Common analytical and interpretational errors in educational research*. Paper presented at the annual conference of the American Educational Research Association (AERA), New Orleans.

Onwuegbuzie, A.J., & Daniel, L.G. (in press). Uses and misuses of the correlation coefficient. *Research in the Schools*.

Onwuegbuzie, A.J., &Teddlie, C. (in press). Advanced data analysis techniques in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

Outhwaite, W. (1975). *Understanding social life: The method called verstehen*. London: Allen and Unwin.

Outhwaite, W. (1983). *Concept formation in social science*. London: Routledge and Kegan Paul.

Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Reichardt, C.S., & Rallis, S.F.(1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C.S. Reichardt & S.F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 85-92). San Francisco: Jossey-Bass.

Rist, R.C. (1977). On the relations among educational research paradigms: From

disdain to detente. *Anthropology and Education Quarterly, 8*, 42-49.

Sechrest, L., & Sidana, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning, 18*, 77-87.

Smith, J.K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher, 12*, 6-13.

Smith, J.K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher, 15*, 4-13.

Stephenson, W. (1953). *The study of behavior*. Chicago: University of Chicago Press.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Applied Social Research Methods Series (Vol. 46). Thousand Oaks, CA: Sage.

Thompson, B. (1998a, April). *Five methodological errors in educational research: The pantheon of statistical significance and other faux pas*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Thompson, B. (1998b). Statistical testing and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*, 33-38.

Thompson, B. (1999, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Invited address presented at the annual meeting of the American Educational Research Association, Montreal [On-line]. Available: *http://acs.tamu.edu/~bbt6147/aeraad99.htm*

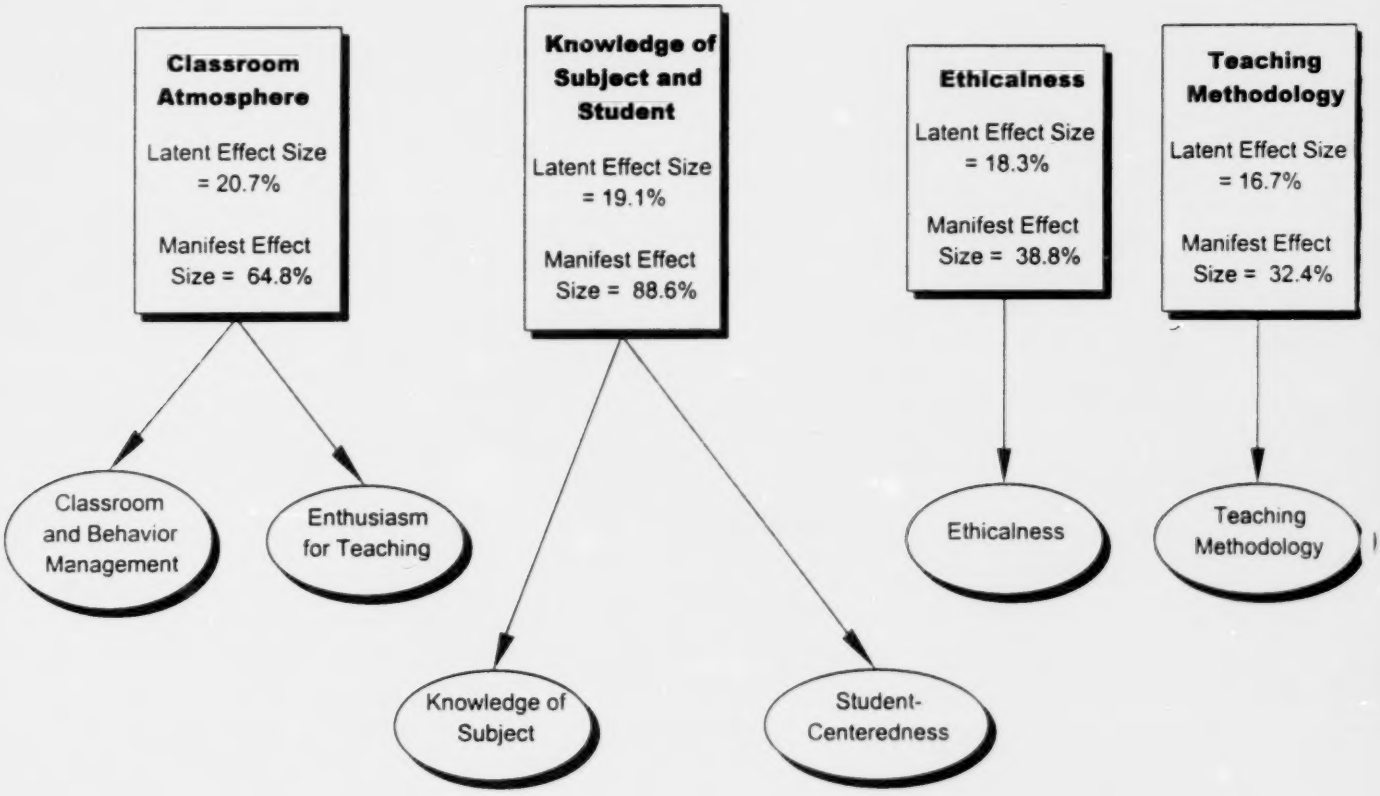
Thompson, B., & Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Witcher, A., Onwuegbuzie, A.J., & Minor, L. (in press). Characteristics of effective teachers: Perceptions of preservice teachers. *Research in the Schools*.

Figure Caption

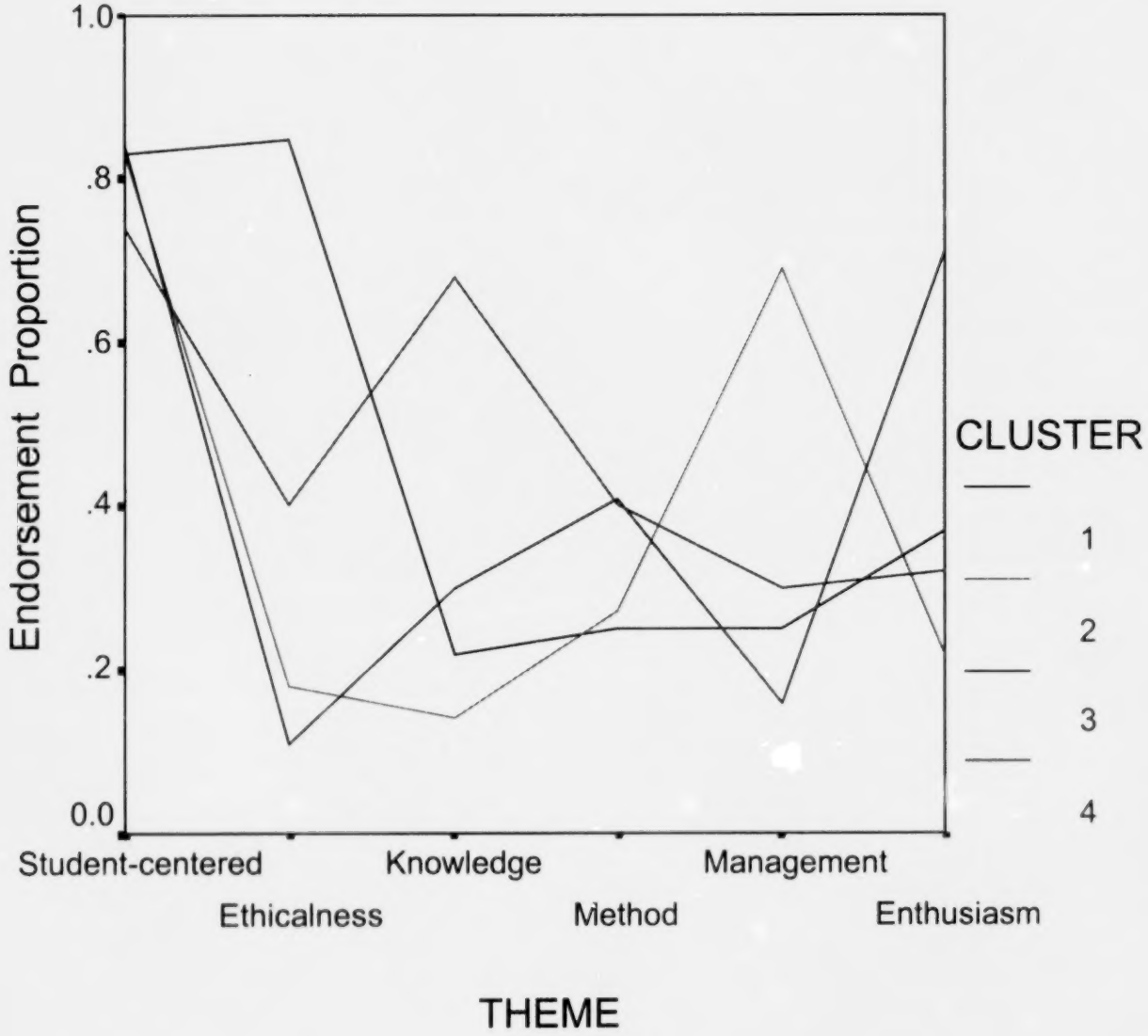
Figure 1. Thematic structure pertaining to preservice teachers' perceptions of the characteristics of effective teachers.



BEST COPY AVAILABLE

Figure Caption

Figure 2. Average profiles Relating to preservice teachers' perceptions of the characteristics of effective teachers.


1.0
.8
.6
.4
.2
0.0
Endorsement Proportion
CLUSTER
1
2
3
4
Student-centered
Ethicalness
Knowledge
Method
Management
Enthusiasm
THEME